View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Computers and Mathematics with Applications 64 (2012) 1100–1107

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis

Chen-Chung Liu [a], Chung-Yen Tsai [b], Jui Liu [c], Chun-Yuan Yu [b], Shyr-Shen Yu [b,*]

[a] Department of Electronic Engineering, National Chin-Yi University of Technology, Taichung 411, Taiwan

[b] Department of Computer Science and Engineering, National Chung-Hsing University, Taichung 400, Taiwan

[c] Department of General Medicine, Royal North Shore Hospital, Sydney, Australia

ARTICLE INFO

Keywords:
Mammogram
Pectoral muscle
Otsu thresholding
Multiple regression analysis (MRA)

ABSTRACT

One of the issues when interpreting a mammogram is that the density of a pectoral muscle region is similar to the tumor cells. The appearance of pectoral muscle on medio-lateral oblique (MLO) views of mammograms will increase the false positives in computer aided detection (CAD) of breast cancer. For this reason, pectoral muscle has to be identified and segmented from the breast region in a mammogram before further analysis. The main goal of this paper is to propose an accurate and efficient algorithm of pectoral muscle extraction on MLO mammograms. The proposed algorithm is based on the positional characteristic of pectoral muscle in a breast region to combine the iterative Otsu thresholding scheme and the mathematic morphological processing to find a rough border of the pectoral muscle. The multiple regression analysis (MRA) is then employed on this rough border to obtain an accurate segmentation of the pectoral muscle. The presented algorithm is tested on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. The experimental results show that the pectoral muscle extracted by the presented algorithm approximately follows that extracted by an expert radiologist.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Breast cancer is the most common cancer in women — which constitutes approximately 25% of all cancers in women, and every one in eight women develop breast cancer. Successful treatment relies on early detection, and screening mammography targeting high risk groups is a common practice in developed countries.

Extracting the breast region accurately from a mammogram is a kernel stage for mammography. It significantly influences the overall analysis accuracy and processing speed of the whole breast mass analysis. It is due to the density of a pectoral muscle region being similar to that of the mammographic parenchyma (tumor). Many studies on tumor detection on a mammogram have shown that the appearance of pectoral muscle in medio-lateral oblique (MLO) views of mammograms will increase the false positive in computer aided detection (CAD) of breast cancer [1–3]. Therefore, successful identification and segmentation of pectoral muscle from the breast region on a mammogram before further analysis should improve the accuracy when interpreting the mammogram.

Only a few studies have been presented in the literature to address the pectoral muscle detection. Nagi et al. [4] used morphological preprocessing and seeded region growing to detect the pectoral muscle. Yapa et al. [5] segment the pectoral muscle region by utilizing the combination of an improved fast-marching method and mathematical morphological

* Corresponding author.
E-mail addresses: ccl@ncut.edu.tw (C.-C. Liu), james1022jk@gmail.com (C.-Y. Tsai), titoakl@yahoo.com (J. Liu), t112@nkut.edu.tw (C.-Y. Yu), epyu@nchu.edu.tw (S.-S. Yu).

0898-1221/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2012.03.028

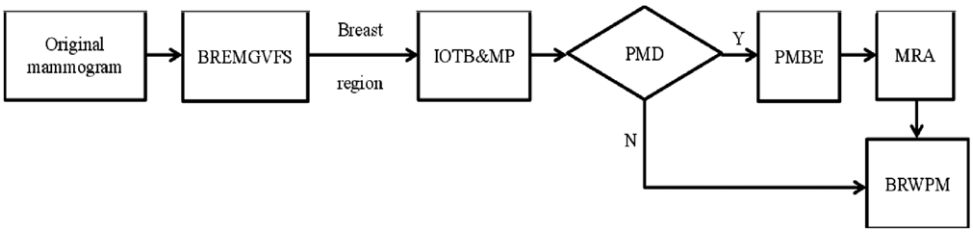

Fig. 1. The flow chart of the proposed pectoral muscle segmentation algorithm for mammograms.

operators such as area morphology, alternating sequential filter, openings and closings. Kwok et al. [6] detected the pectoral muscle edge in a restricted neighborhood derived by estimating the pectoral muscle edge as a straight line segment, and then refined the detected edge by surface smoothing. However, the hypothesis of a straight line segment for the representation of the pectoral muscle edge is not always correct.

Multiple regression analysis (MRA) is a statistical method used to model the relationships between several independent variables and a dependent variable by fitting an equation to the observed data [7,8]. Numerous experiments have shown that multiple regression models can be used to make accurate predictions. Multiple regression procedures are very widely used in the social and natural sciences today [9]. It is also a suitable technique for us to modify the extracted pectoral muscle border to increasing the precision of the proposed pectoral muscle segmentation algorithm.

The proposed pectoral muscle segmentation algorithm bases on the positional characteristic of pectoral muscle in a breast region to combine the iterative Otsu thresholding scheme and the mathematic morphological processing to find a rough border of the pectoral muscle. Once the pectoral muscle border is extracted, the multiple regression analysis is then employed on the rough border to obtain an accurate segmentation of the pectoral muscle. The presented algorithm is tested on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. The experimental results show that (i) the pectoral muscle extracted by the presented algorithm approximately follows that extracted by an expert radiologist, (ii) the proposed scheme is adaptive to large variations in appearance of the pectoral muscle; it remains effective when the pectoral border is obscured by superimposed muscle tissue or artifacts, (iii) the proposed scheme will decrease the false positives in computer aided detection (CAD) of breast cancer. The remainder of this paper is organized in the following order: Section 2 introduces the presented extraction algorithm of pectoral muscle region from mammograms. Section 3 presents the experimental results. The conclusions of this paper are presented in Section 4.

2. Proposed pectoral muscle segmentation algorithm

In order to construct an accurate pectoral muscle segmentation algorithm for digital mammograms, several schemes are used in this paper. The overall pectoral muscle segmentation algorithm for digital mammograms is shown in Fig. 1. There are mainly three stages involved in the algorithm: (i) breast region extracted by using modified gradient vector flow (MGVF) snake, (ii) rough pectoral muscle region detection by using the iterative Otsu thresholding scheme, (iii) pectoral border modification by utilizing multiple regression analysis. The details of these stages used in the presented pectoral muscle segmentation algorithm are described in the following subsections.

2.1. Breast region extracted by using modified gradient vector flow (MGVF) snake [10]

To segment a digital mammogram into the breast region and the background is an essential step for pectoral muscle segmentation. By extracting the breast region, the precision of the pectoral muscle segmentation is increased and the time consumption is shortened. The breast region extraction algorithm must be fully automated and give correct results for all digital mammograms. The breast region segmentation by using an MGVF snake is employed in this paper, a method that was introduced in our previous paper: "*a breast extraction scheme for digital mammograms using gradient vector flow snake*" [10]. The proposed breast region extracting algorithm integrated the median filtering step, the scale down step, the binarization processing step, the morphological erosion processing step and novel gradient adjusting step. The median filter step was used to filter out the noise in a mammogram, the scale down step was used to resize down the mammogram size to speed up the breast region extraction. The binarization processing step and the morphological erosion processing step were used to find a rough breast border. The novel gradient adjusting step was applied to get a modified edge map; and then the gradient vector flow snake (GVF snake) was employed to obtain an accurate breast border from the rough breast border. This breast extraction scheme can output the corresponding accurate breast region from an input digit mammogram, and the output breast region will be used in the following stage to detect a rough border of the pectoral muscle.

2.2. Detection of pectoral muscle rough border

For obtaining a better initial extraction, the binarization processing stage uses the Otsu thresholding scheme to binarize the extracted breast region into a black–white image iteratively. The binarization procedure shall be terminated while the absolute difference ratio of the white area between two adjacency iterations is less than 5%. In addition, if the white area locates either at the upper left or upper right corner of the breast region, the white region is taken as the rough pectoral muscle region. Finally, the morphological erosion processing is applied on the rough segmented pectoral muscle area to find a clearer border of the pectoral muscle area.

2.2.1. Binarization

The binarization step is used to obtain a rough pectoral muscle area from the breast region. On the other hand, the gray value of the real border of the pectoral muscle is similar to that of the background in a grayscale mammogram. The presented algorithm repeats the famous Otsu thresholding scheme to binarize the breast region to determine the rough pectoral muscle area. The Otsu thresholding scheme proposed by Otsu (1979) searches an optimal threshold to divide a grayscale image's pixels into two classes. The optimal threshold is evaluated by the discriminated criterion which maximizes the separability between target and background classes. The Otsu thresholding scheme inputs a data set and determines the maximum and minimum values of the input data set, indicated as $L_{\min}$ and $L_{\max}$. The histogram of the data set is normalized as a probability distribution by the following equation.

$$p(l) = n(l)/N, \quad p(l) \geq 0, \qquad \int_{l=L_{\min}}^{L_{\max}} p(l)^* dl = 1. \tag{1}$$

Here, $n(l)$ is the number of elements with value l and N is the total number of elements of the data set. We suppose that all elements of the set are divided into two classes, C_1 and C_2 by a threshold k. Then, the probabilities of occurrence, ω, and mean, μ, of each class are evaluated by the following formulas.

$$\omega_1 = \Pr(C_1) = \frac{1}{N} \int_{l=L_{\min}}^{k} n(l)^* dl = \omega(k), \tag{2}$$

$$\omega_2 = \Pr(C_2) = \frac{1}{N} \int_{l=k}^{L_{\max}} n(l)^* dl = 1 - \omega_1 = 1 - \omega(k), \tag{3}$$

$$\mu_1 = \int_{l=L_{\min}}^{k} l^* \Pr(l|C_1)^* dl = \int_{l=L_{\min}}^{k} l^* p(l)^* dl/\omega_1 = \mu(k)/\omega(k), \tag{4}$$

$$\mu_2 = \int_{l=k}^{L_{\max}} l^* \Pr(l|C_2)^* dl = \int_{l=k}^{L_{\max}} l^* p(l)^* dl/\omega_2 = (\mu_T - \mu(k))/(1 - \omega(k)), \tag{5}$$

where

$$\mu_T = \mu(L_{\max}) = \int_{l=L_{\min}}^{L_{\max}} l^* p(l)^* dl. \tag{6}$$

The class variances are evaluated by

$$\sigma_1^2 = \int_{l=L_{\min}}^{k} (l - \mu_1)^{2*} \Pr(l|C_1)^* dl = \int_{l=L_{\min}}^{k} (l - \mu_1)^{2*} p(l)^* dl/\omega_1, \tag{7}$$

$$\sigma_2^2 = \int_{l=k}^{L_{\max}} (l - \mu_2)^{2*} \Pr(l|C_2)^* dl = \int_{l=k}^{L_{\max}} (l - \mu_2)^{2*} p(l)^* dl/\omega_2. \tag{8}$$

The within-class variance, the between-class variance and the total variance of element-values are defined as follows.

$$\sigma_W^2 = \omega_1^* \sigma_1^2 + \omega_2^* \sigma_2^2, \tag{9}$$

$$\sigma_B^2 = \omega_1(\mu_1 - \mu_T)^2 + \omega_2(\mu_2 - \mu_T)^2 = \omega_1^* \omega_2(\mu_1 - \mu_2)^2, \tag{10}$$

$$\sigma_T^2 = \sigma_W^2 + \sigma_B^2 = \int_{l=L_{\min}}^{L_{\max}} (l - \mu_T)^{2*} p(l)^* dl. \tag{11}$$

Otsu introduced the following two measuring functions to get the optimal threshold $\hat{k}$:

$$\hat{k} = \underset{L_{\min} \leq k \leq L_{\max}}{\arg\max} \ (\sigma_B^2(k)) = \underset{L_{\min} \leq k \leq L_{\max}}{\arg\max} \ (\omega_1^* \omega_2(\mu_1 - \mu_2)^2), \tag{12}$$

$$\hat{k} = \underset{L_{\min} \leq k \leq L_{\max}}{\arg\min} \ (\sigma_W^2(k)) = \underset{L_{\min} \leq k \leq L_{\max}}{\arg\min} \ (\omega_1^* \sigma_1^2 + \omega_2^* \sigma_2^2). \tag{13}$$

In this paper, the black region in a mammogram is the background, and white regions usually contained the pectoral muscle. For obtaining a region as the pectoral muscle, the presented algorithm will apply the morphological operation and connected component scheme on the white regions colored by the Otsu thresholding scheme to extract the pectoral muscle area in following processes.

2.2.2. Morphological processing

In an Otsu binarized mammogram, the white area located either at the left upper or right upper corner of the breast region may be the pectoral muscle. For obtaining a better and simple binary image of pectoral muscle, the presented algorithm adapts the morphological opening operation to delete the non-pectoral-muscle white areas. The structuring element of the morphological opening operation used in the presented algorithm is a disk with radius 2 pixels.

2.2.3. Extract pectoral muscle border

Many non-pectoral-muscle white areas still remain in the binary mammograms that have been operated by the morphological opening operation, but they don't locate at either the left upper or right upper corner of the breast regions. The presented algorithm utilizes the location characteristic of the pectoral muscle in a breast region to extract the upper left or upper right corner white area as the pectoral muscle area I. Those white areas located outside the left upper and right upper corners will be set to be the background IB, and the gray levels of IB are reset as zero (black). The initial pectoral muscle border is extracted by subtracting the eroded IB from the IB before eroding.

2.3. The modification of pectoral muscle border using multiple regression analysis [9]

When performing experiments, data are frequently tabulated in the form of ordered pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ with each x_i distinct. Given the data, it is then usually desirable to be able to predict y from x by finding a mathematical model, that is, a function $y = H(x)$ that fits the data as closely as possible. One way to determine how well the function $y = H(x)$ fits these order pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ is to measure the sum of squares of the errors (SSE) between the predicted values of y and the observed values y_i for all of the n data points.

Multiple regression is one of the widely used statistical techniques [11]. This technique is used to find a polynomial function of degree k, $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k$ as the predicting function, that has the minimum of the sum of squares of the errors(SSE) between the predicted values of y and the observed values y_i for all of the n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. The values of $\beta_0, \beta_1, \beta_2, \ldots,$ and β_k that minimize

$$SSE(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k)]^2, \tag{14}$$

are obtained by setting the $k + 1$ first partial derivatives $\frac{\partial}{\partial \beta_0} SSE(\beta_0, \beta_1, \ldots, \beta_k)$, $\frac{\partial}{\partial \beta_1} SSE(\beta_0, \beta_1, \ldots, \beta_k), \ldots,$ and $\frac{\partial}{\partial \beta_k} SSE(\beta_0, \beta_1, \ldots, \beta_k)$ equal to zero, and solving the resulting simultaneous linear system of the so-called normal equations:

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^2 + \cdots + \beta_k \sum_{i=1}^{n} x_i^k = \sum_{i=1}^{n} y_i, \tag{15}$$

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 + \beta_2 \sum_{i=1}^{n} x_i^3 + \cdots + \beta_k \sum_{i=1}^{n} x_i^{k+1} = \sum_{i=1}^{n} x_i y_i, \tag{16}$$

$$\beta_0 \sum_{i=1}^{n} x_i^k + \beta_1 \sum_{i=1}^{n} x_i^{k+1} + \beta_2 \sum_{i=1}^{n} x_i^{k+2} + \cdots + \beta_k \sum_{i=1}^{n} x_i^{2k} = \sum_{i=1}^{n} x_i^k y_i, \tag{17}$$

and the matrix form solution of the normal equations system is

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \mathrm{B} = \left[X^T X\right]^{-1} \left[X^T Y\right], \tag{18}$$

where

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ 1 & x_3 & x_3^2 & \cdots & x_3^k \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix}, \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \tag{19}$$

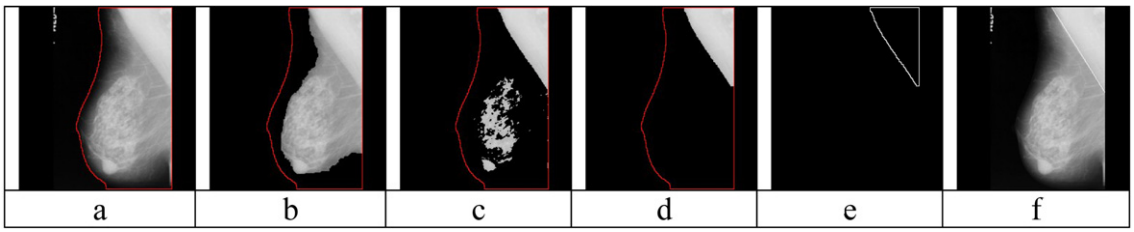

Fig. 2. The output of each steps of the presented algorithm, (a) breast region extracted by using MGVF snake, (b) after Otsu thresholding and morphological processing one time, (c) after Otsu thresholding and morphological processing two times, (d) after position detecting, (e) extracted pectoral muscle rough border, (f) final extracted pectoral muscle.

The multiple regression analysis is then applied to the pectoral muscle border obtained in the previous section to determine the more precise pectoral muscle border. Fig. 2 shows the output of each step of the presented algorithm.

3. Experimental results

The presented pectoral muscle segmentation scheme was applied to the Mammogram Image Analysis Society (MIAS) database. This MIAS database has 322 mammograms and each mammogram is 50 um/pixel. The size of a mammogram is 1024*1024 with a bit-depth of 8 bits ([0, 255]). All the 322 mammograms have been tested to find that 150 mammograms have pectoral muscle and 172 mammograms do not have any pectoral muscle. Each of these 150 pectoral muscles including mammograms are labeled by an expert radiologist with hand carefully to construct the ground truth extraction for evaluating the presented algorithm's extractions.

Segmentation results are analyzed according the normalized performance metrics rather than the number of pixels of the areas between the true and computed boundaries to show the different performance features of the segmentation algorithms. These normalized performance metrics are mean error (ME1) function, misclassification error (ME2) function, relative foreground area error (RFAE), extraction error rate (EER), region non-uniformity (NU), and modified Hausdorff distance (MHD) [12,13,11]. ME1, ME2, RFAE, EER and MHD are varying from 0 for a perfectly correct segmentation to 1 for a completely error case. These performance measures are illustrated as follows:

(i) The $ME1$ function is defined as:

$$ME1 = \sum_{i=1}^{n} \sqrt{(x\_m_i - x_i)^2 + (y\_m_i - y_i)^2}/n, \tag{20}$$

where n is total number of pectoral muscle border points, $(x\_m_i,\ y\_m_i)$ is the Cartesian coordinate of the i-th point of the manual border of the pectoral muscle, and (x_i, y_i) is the Cartesian coordinate of the detected pectoral muscle point which is closest to the i-th point of the manual border of the pectoral muscle. If the manual border and the detected border are closer, the value of $ME1$ is closer to zero.

(ii) The misclassification error ($ME2$) evaluates the inaccuracy of an algorithm which is defined as:

$$ME2 = 1 - (TP + TN)/(TP + FN + TN + TP) = (FN + FP)/(TP + FN + TN + TP), \tag{21}$$

where TP, TN, FP, and FN represent the areas of true positive, true negative, false positive and false negative, respectively.

(iii) The relative foreground area error ($RFAE$) evaluates the region mismatch between the extracted object and the ground-truth object and is defined as:

$$RFAE = \begin{cases} \dfrac{(TP + FN) - (FP + TP)}{TP + FN} = \dfrac{FN - FP}{TP + FN} & \text{if } (FP + TP) < (TP + FN) \\ \dfrac{(FP + FN) - (TP + FN)}{FP + TP} = \dfrac{FP - FN}{FP + TP} & \text{if } (FP + TP) \geq (TP + FN) \end{cases}, \tag{22}$$

where $FP + TP$ indicates the extracted object while $TP + FN$ denotes the ground-truth object.

(iv) The extraction error rate (EER) evaluates the failure rate of the algorithm. It is the summation of under-extraction rate (UER) and the over-extraction (OER) rate and is defined as:

$$EER = UER + OER = FN/(TP + FN) + FP/(TP + FN). \tag{23}$$

(v) The region non-uniformity (NU) evaluates the distinguishable quality between the extracted objects and background. It does not require a ground-truth image and is defined as:

$$NU = \frac{TP + FP}{(TP + FP) + (TN + FN)} \times \frac{\sigma_e^2}{\sigma^2}, \tag{24}$$

where σ_e^2 denotes the variance of the extracted object and σ^2 is the variance of the original image.

Table 1
The MOEs of the pectoral muscle border of mdb005 obtained by MRA with a polynomial of degree n, $n = 1, 2, \ldots, 10$.

Poly. Degree	1	2	3	4	5	6	7	8	9	10
ME1	0.3411	0.2957	0.2062	0.1812	0.1803	0.1803	0.1803	0.1803	0.1803	0.1803
ME2	0.0371	0.0335	0.0219	0.0179	0.0173	0.0172	0.0172	0.0172	0.0172	0.0172
RFAE	0.0042	0.0038	0.0027	0.0020	0.0020	0.0020	0.0019	0.0019	0.0019	0.0019
EER	0.0130	0.0097	0.0085	0.0079	0.0076	0.0077	0.0077	0.0076	0.0076	0.0076
NU	0.0258	0.0224	0.0186	0.0152	0.0148	0.0149	0.0148	0.0148	0.0148	0.0148
MHD	0.3018	0.2503	0.1831	0.1456	0.1410	0.1410	0.1410	0.1410	0.1409	0.1409

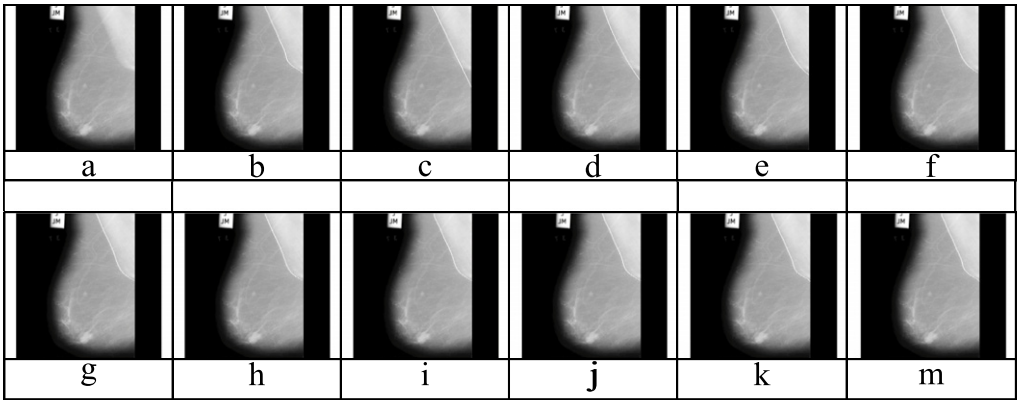

Fig. 3. The pectoral muscle border of mdb005 constructed by MRA with different degree polynomials; (a) original mammogram, (b) ground truth labeled by an expert radiologist, (c) degree 1, (d) degree 2, (e) degree 3, (f) degree 4, (g) degree 5, (h) degree 6, (i) degree 7, (j) degree 8, (k) degree 9, (m) degree 10.

(vi) The modified Hausdorff distance (*MHD*) is utilized to assess the shape similarity of the extracted object O_e to the ground-truth object O_g, which is defined as:

$$MHD(O_e, O_g) = \frac{1}{|O_g|} \sum_{p \in O_g} d(p, O_e), \tag{25}$$

where $|O_g|$ is the pixel number of the ground-truth object O_g, $d(p, O_e)$ is the distance from pixel p to the extracted object O_e which is defined as:

$$d(p, O_e) = \min\{\text{distance}(p, q) | q \in O_e\}, \tag{26}$$

where distance(p, q) is the Euclidean distance between pixels p and q.

Fig. 3 shows the pectoral muscle borders of mdb005 constructed by MRA polynomials with different degrees to demonstrate the performance in vision of the variation of the degrees of the MRA polynomials. Table 1 shows the corresponding MOEs of Fig. 3. Table 1 illustrates that the pectoral muscle border constructed with degree one polynomial for mdb005 is the worst, and the pectoral muscle border constructed with degree ten polynomial for mdb005 is the best. Although both Fig. 3 and Table 1 show that the higher degree polynomial fitting functions have higher accuracy for pectoral muscle border detection, they also show that the precisions of the pectoral muscle borders fitted by polynomials with degrees more than four are almost the same. Hence, the degree five polynomial is taken as the optimal polynomial for the proposed pectoral muscle segmentation not only to reduce the complexity and the time consumption, but also to obtain the highly accurate segmentation.

To demonstrate the better performance of the proposed algorithm compared to the other related schemes, the scheme of Ferrar et al. [14] is used as a comparison. The performance comparison between the proposed and the Ferrar et al. schemes for the segmentation results for the 322 mammograms of the Mammogram Image Analysis Society (MIAS) database are illustrated in Fig. 4, Tables 2 and 3.

Fig. 4 is given to compare the visual results obtained by using Ferrar et al. and the proposed schemes on six pectoral muscles including mammograms. In Fig. 4, row 1 shows the visualizations of six original mammograms for experiment, row 2 shows the corresponding ground truth of pectoral muscle labeled by an expert radiologist with hand carefully. Rows 3 and 4 show the corresponding segmentation results obtained by using the method of Ferrar et al. and the proposed algorithm, respectively.

The pectoral muscle may appear on the mammogram as very small or very large, as convex or concave with different curvatures as shown in Fig. 4(a)–(e). They were all segmented accurately to show that the proposed algorithm is adaptive to variations in pectoral muscle size, density and curvature. Fig. 4(a)–(d) show that although the pectoral muscles are obscured by tapes of different thickness, they were all segmented successfully. The results illustrate that the proposed segmentation

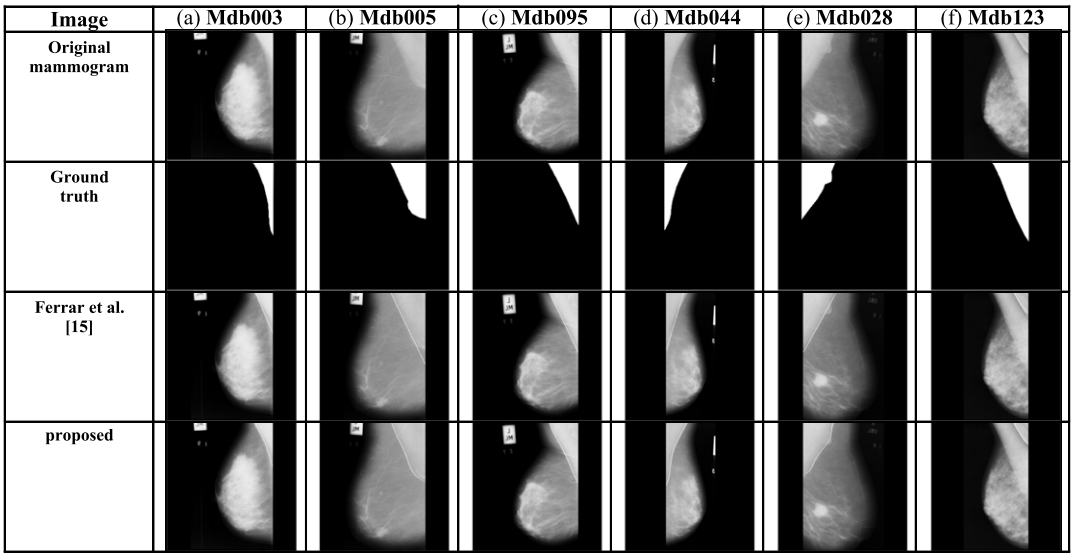

Fig. 4. The visual results obtained by using Ferrar et al. and the proposed schemes on six pectoral muscle including mammograms of the Mammogram Image Analysis Society (MIAS) database.

Table 2
The corresponding six MOEs of the segmentation results obtained by using Ferrar et al. and the proposed schemes on the six pectoral muscle including mammograms in Fig. 4.

Mammogram	ME1		ME2		RFAE		EER		NU		MHD	
	Ferrar et al. [14]	Proposed	Ferrar et al. [14]	Proposed	Ferrar et al. [14]	Proposed	Ferrar et al. [14]	Proposed	Ferrar et al. [14]	Proposed	Ferrar et al. [14]	Proposed
Mdb003	1.0109	0.8582	0.4271	0.4036	0.0766	0.0690	0.0597	0.0597	0.1249	0.1249	0.8425	0.8425
Mdb005	0.3411	0.1803	0.0371	0.0173	0.0042	0.0020	0.0130	0.0076	0.0258	0.0148	0.3018	0.1410
Mdb028	1.6287	0.7023	0.2248	0.0847	0.0855	0.0086	0.0270	0.0061	0.0184	0.0042	1.5336	0.5218
Mdb044	0.9068	0.6801	0.3856	0.0814	0.0647	0.0063	0.0472	0.0050	0.0936	0.0038	0.5369	0.5574
Mdb095	0.1303	0.1141	0.0016	0.0015	0.0000	0.0000	0.0050	0.0048	0.0002	0.0002	0.1273	0.0983
Mdb123	20.958	20.695	0.7880	0.7169	0.0852	0.0784	0.5709	0.5206	0.1427	0.1382	12.977	12.558

Table 3
The averages and the ranges (minimum–maximum) of MOEs for the 150 pectoral muscle including mammograms of the MIAS.

Scheme	ME1	ME2	RFAE	EER	NU	MHD
Ferrar et al. [14]	2.2103 (0.1075–23.6711)	0.0114 (0.0013–0.7545)	0.0110 (0.0000–0.1537)	0.0299 (0.0045–0.1225)	0.0051 (0.0002–0.1696)	2.0088 (0.1835–14.7111)
Proposed	1.7188 (0.1102–22.6817)	0.0083 (0.0011–0.7380)	0.0056 (0.0000–0.1281)	0.0134 (0.0027–0.1092)	0.0014 (0.0001–0.1688)	0.8702 (0.0858–13.7255)

algorithm is robust against artifacts such as sticky tapes. Sometimes there may be several sharp corners on the boundary of the pectoral muscle region [see Fig. 4(e)] to cause poor segmentation of low degree MRA. This case can be easily improved by increasing the degree of MRA with little processing time cost. However, in a small number of cases, there may be more than one layer of muscle tissue in the pectoral region and more than one edge may be detected. In such cases the detected curve is more likely to be placed at the upper edge rather than at the lower edge. The reasons for this are that (i) the initial pectoral muscle border extraction scheme uses histogram thresholding instead of edge detection to find the edge; and (ii) the initial pectoral muscle region grows from the upper right or left corner area by merging its neighborhood pixels. Therefore, the upper edge is more likely to be extracted. Fig. 4(f) shows pectoral muscle with two edges of different edge strengths and the upper edge is extracted as the pectoral muscle border. The axillary muscle tissue is not normally seen on the MLO view, and it is frequently the result of poor positioning.

The corresponding six MOEs of the segmentation results obtained by using Ferrar et al. and the proposed schemes on six pectoral muscles including mammograms are summarized in Table 2. From Table 2, the proposed algorithm achieves smaller MOEs than the method of Ferrar et al.. Both Fig. 4 and Table 2 clearly show that the Otsu thresholding scheme together with multiple regression analysis scheme provides a much better result than the method of Ferrar et al. for segmenting the pectoral muscle from breast regions.

The averages and the ranges (minimum–maximum) of MOEs for the 150 pectoral muscles including mammograms of the MIAS database are given in Table 3. Table 3 shows that all the averages and the ranges of the six MOEs of the method

of Ferrar et al. are much larger than that for the proposed scheme. These facts indicate that the proposed method provides more consistent results than the method of Ferrar et al.. The method of Ferrar et al. is very accurate for a straight line segment pectoral muscle border. Since many mammograms' pectoral muscle edge curves away from the chest wall toward the middle and lower-half of the images, the segmentation accuracy of pectoral muscle in the method of Ferrar et al. degrades significantly while a mammogram pectoral muscle border is not a straight line segment. The proposed method takes an additional step, the multiple regression analysis step, to modify the pectoral muscle border extracted by the iterative Otsu thresholding scheme. Compared to the method of Ferrar et al., the accuracy is much higher in the proposed method. Table 3 shows that the proposed scheme can give precise extractions for pectoral muscles from mammograms.

4. Conclusions

Breast cancer is the most common cancer in women — which constitutes approximately 25% of all cancers in women, and every one in eight women develop breast cancer. Successful treatment relies on early detection, and screening mammography targeting high risk groups is a common practice in developed countries. Extracting the breast region accurately from a mammogram is a kernel stage for mammography. The appearance of pectoral muscle in medio-lateral oblique (MLO) views of mammograms can increase the false positives in the computer aided detection (CAD) of breast cancer. For this reason, pectoral muscle has to be identified and segmented from the breast region in a mammogram before further analysis. We have proposed an accurate and efficient algorithm to automatically segment the pectoral muscle in MLO mammograms. The proposed algorithm is based on the positional characteristic of pectoral muscle in a breast region and combines the iterative Otsu thresholding scheme and mathematic morphological processing to find the rough border of the pectoral muscle. The multiple regression analysis is then employed to overcome the limitation of the straight-line hypothesis imposed by other algorithms to obtain the accurate segmentation of the pectoral muscle. The presented algorithm is tested on 150 pectoral muscles including digital mammograms from the Mammogram Image Analysis Society (MIAS) database. In the experiments, several measures of errors, such as mean error (ME1) function, misclassification error (ME2) function, relative foreground area error (RFAE), extraction error rate (EER), region non-uniformity (NU), and modified Hausdorff distance (MHD) were conducted to measure the performance of the proposed algorithm. With reference to the manually demarcated pectoral muscle regions, the segmented regions provided by the proposed scheme resulted in low average ME1, ME2, RFAE, EER and MHD with 1.7188, 0.0083, 0.0056, 0.0134 and 0.8702, respectively. The experimental results show that the pectoral muscle extracted by the presented algorithm approximately follows that extracted by an expert radiologist. The experimental results also show that the proposed scheme is adaptive to large variations in the appearance of the pectoral muscle; it remains effective when the pectoral border is obscured by superimposed muscle tissue or artifacts. In the future we will develop a high performance breast mass analysis based on accurate breast region segmentation to power the computer aided detection of breast cancer.

Acknowledgments

This research was supported by the National Science Council R.O.C. under Grant NSC 100-2221-E-005-086 and NSC 100-2221-E-167-028.

References

[1] G. Kom, A. Tiedeu, M. Kom, Automated detection of masses in mammograms by local adaptive thresholding, Computers in Biology and Medicine 37 (2007) 37–48.
[2] P. Delogu, M.E. Fantacci, P. Kasae, A. Reticoa, Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier, Computers in Biology and Medicine 37 (2007) 1479–1491.
[3] A.R. Domiınguez, A.K. Nandi, Detection of masses in mammograms via statistically based enhancement, multilevel thresholding segmentation, and region selection, Computerized Medical Imaging and Graphics 32 (2008) 304–315.
[4] J. Nagi, S.A. Kareem, F. Nagi, S.K. Ahmed, Automated breast profile segmentation for ROI detection using digital mammograms, in: 2010 IEEE EMBS Conference on Biomedical Engineering & Sciences, 2010, pp. 87–92.
[5] R.D. Yapa, K. Harada, Breast skin-line estimation and breast segmentation in mammograms using fast-marching method, International Journal of Biomedical Sciences 3 (1) (2008) 54–62.
[6] S.M. Kwok, R. Chandrasekhar, Y. Attikiouzel, M.T. Rickard, Automatic pectoral muscle segmentation on mediolateral oblique view mammograms, IEEE Transactions on Medical Imaging 23 (9) (2004) 1129–1140.
[7] Y.L. Feng, S.G. Lv, Unified approach to coefficient-based regularized regression, Computers and Mathematics with Applications 62 (2011) 506–515.
[8] V. Strijov, G.W. Weberb, Nonlinear regression model generation using hyperparameter optimization, Computers and Mathematics with Applications 60 (2010) 981–988.
[9] C.C. Liu, W.Y. Chen, Screw pitch precision measurement using simple linear regression and image analysis, Applied Mathematics and Computation 178 (2006) 390–404.
[10] S.S. Yu, C.Y. Tsai, C.C. Liu, A breast extraction scheme for digital mammograms using gradient vector flow snake, in: 4th International Conference on New Trends in Information Science and Service Science, Gyeongju, Korea, 2010, pp. 715–720.
[11] K. Qin, K. Xua, F. Liu, D. Li, Image segmentation based on histogram analysis utilizing the cloud model, Computers and Mathematics with Applications 62 (2011) 2824–2833.
[12] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging 13 (1) (2004) 146–165.
[13] M.H. Horng, Performance evaluation of multiple classification of the ultrasonic supraspinatus images by using ML, RBFNN and SVM classifiers, Expert Systems with Applications 37 (2010) 4146–4155.
[14] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, A.F. Frère, Automatic identification of the pectoral muscle in mammograms, IEEE Transactions on Medical Imaging 23 (2) (2004) 232–245.